

Automatic Alignment of Indoor and Outdoor Building Models using 3D Line Segments

Tobias Koch, Marco Körner
Remote Sensing Technology
Technical University of Munich
{tobias.koch,marco.koerner}@tum.de

Friedrich Fraundorfer
Institute for Computer Graphics and Vision
Graz University of Technology
fraundorfer@icg.tugraz.at

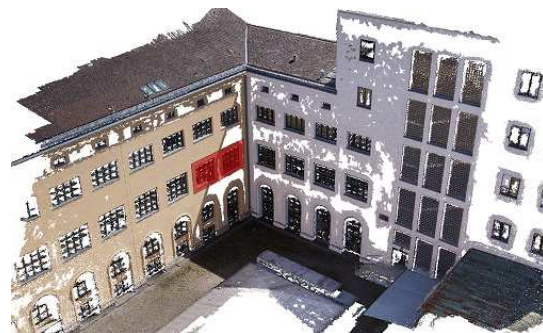
Abstract

This paper presents an approach for automatically aligning the non-overlapping interior and exterior parts of a 3D building model computed from image based 3D reconstructions. We propose a method to align the 3D reconstructions by identifying corresponding 3D structures that are part of the interior and exterior model (e.g. openings like windows). In this context, we point out the potential of using 3D line segments to enrich the information of point clouds generated by SfMs and show how this can be used for interpreting the scene and matching individual reconstructions.

1. Introduction

A cheap and fast way for generating building models is to obtain 3D information from image sequences. Typically, 3D reconstruction pipelines like *Structure-from-Motion* (SfM) followed by *Multi-View Stereo* (MVS), and meshing are used for small and large scale reconstructions. With the increasing research on image-based indoor modeling in the recent past, an integration of indoor and outdoor models of the same building is consequently the next step. For instance, Fig. 1 shows the reconstruction of our computer-lab which should be connected to the outdoor façade of the building. When trying to fit a model of the building interior into an existing outdoor model, typically there are no visual correspondences for the alignment using tie points. Therefore, manual work is needed, like using CAD models or floor plans. An automated way providing the true or at least the most probable locations in the outdoor model assumes to reduce human interaction.

Since performing a complete reconstruction using continuous image sequences capturing the entire scene by moving from the outside into the inside of the building is either inaccurate caused by drifts or even unfeasible by the lack of matchable features in most cases, an approach using individual reconstructions is desirable. This also allows for



(a) building exterior



(b) building interior

Figure 1: Dense point clouds of (a) a building façade from images captured by an UAV and (b) inside our computer lab. The true location of the lab is indicated by the red polygon in (a).

matching models generated from image sequences acquired at different points in time.

The most challenging task in matching indoor and outdoor models is to find structures that appear in both image sets but do not describe physically the same part of the scene. To achieve an alignment of these models, topological structures must be found which can be seen from both inside and outside the building, like windows and doors. Identifying

and detecting these objects could be done by semantic image segmentation (scene parsing [5]) or point cloud analysis [10]. Exploiting the fact that window and door frames can be characterized by dominant and co-planar edges, we propose a method employing 3D line segments.

The contribution of this work is (i) a novel framework for aligning individual image-based 3D reconstructions by (ii) using 3D lines for detecting and matching shared geometric structures in different 3D models.

2. Related Work

Although the field of 3D reconstruction, scene interpretation, and modelling of man-made objects like buildings is a well-known and widely studied research topic, there is, to our best knowledge, only little research investigating the question how to automatically align 3D indoor and outdoor models reconstructed by individual image sequences.

However, the existing demand of integrating multiple image-based reconstruction models can be demonstrated by the example of the very recent *Chillon Project* [2], which aimed to fully reconstruct the interior and the exterior of a complex castle in Switzerland. Due to different camera models and acquisition modes (terrestrial and aerial), a fully automatic reconstruction process is not possible. Instead, multiple sub-models were generated and projected in the same reference coordinate system afterwards in a rather manual way by using Ground-Control-Points or selecting tie points in the images by hand. Although the result shows an impressive reconstruction of a complex architectural object, it also demonstrates the extensive manual interaction which is still needed to connect multiple sub-models.

Cohen *et al.* [6] propose a method for merging multiple SfM reconstruction models of a single building which can not be merged due to occlusions or insufficient visual overlap. The approach exploits symmetries and repetitive structures of building façades, as well as semantic reasoning to find reasonable connection points of adjacent models and use them for stitching the models.

In our scenario, we face a similar problem of having no visual overlap when trying to stitch indoor and outdoor models. However, in place of finding connection points, which do not exist in the separated models anyway, we try to find shared geometrical structures that appear in both models like window frames and doors. These shapes can be expressed as edge maps and matched to find suitable connections. When trying to match similar shapes in edge images, *chamfer matching* [4] is widely used, especially in presence of clutter and incompleteness. In our approach, we make use of 3D lines to generate such edge maps which are finally tested for suitable correspondences using chamfer matching.

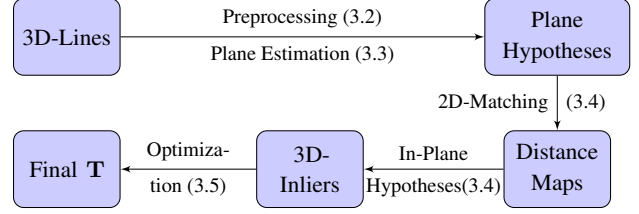


Figure 2: Workflow of the proposed method for aligning building interior and exterior. See denoted chapters for details.

3. System Pipeline

This section describes the pipeline of our proposed model matching approach, as illustrated in Fig. 2. After giving an overview about the basic concept of the method, a detailed description of the individual parts is provided.

3.1. Overview

Given two sets of 3D line segments $L_1 = \{l_1^1, \dots, l_1^n\}$ and $L_2 = \{l_2^1, \dots, l_2^m\}$, the overall goal is to find a transformation $\mathbf{T} = (\mathbf{R}, \mathbf{t}, s)$ to align L_1 to L_2 , where \mathbf{t} , \mathbf{R} and s define the parameters of a 3D similarity transformation as a 3D translation vector, a 3×3 rotation matrix and a scale. Each segment l is defined by its two endpoints. After identifying $i = 1, \dots, k$ corresponding line segments in L_1 and L_2 , the parameters of \mathbf{T} can be estimated by

$$\mathbf{T} = \arg \min_{\mathbf{T}} \sum_{i=1}^k d \left(l_2^i, \pi \left(l_1^i, \hat{\mathbf{T}} \right) \right), \quad (1)$$

where $\pi \left(l, \hat{\mathbf{T}} \right)$ projects a line segment l with $\hat{\mathbf{T}}$, and $d(l_2, l_1)$ computes the length of the perpendicular of two 3D line segments extended to infinity.

As only a small subset out of several thousand pairs of 3D line segments in $L_1 \times L_2$ are expected to be correct 3D line matches, an exhaustive matching scheme is not applicable. Instead, the matching problem is reduced to 2D by defining multiple plane hypotheses in both models, projecting 3D lines onto these planes, and performing 2D binary matching. From the resulting distance maps, local minima can be extracted which indicate potentially matching locations of the indoor model. After coarse alignment and identifying 3D line correspondences, a refinement of \mathbf{T} is applied in 3D by minimizing Eq. (1).

3.2. 3D Line Generation

In a first step, for interior and exterior models, 3D line segments have to be generated from a set of overlapping images. This is realized by initially computing image orientations using, *e.g.*, classical SfM pipelines, like *VSfM* [1], *Pix4D* [3], or *Bundler* [13]. As the following line segment

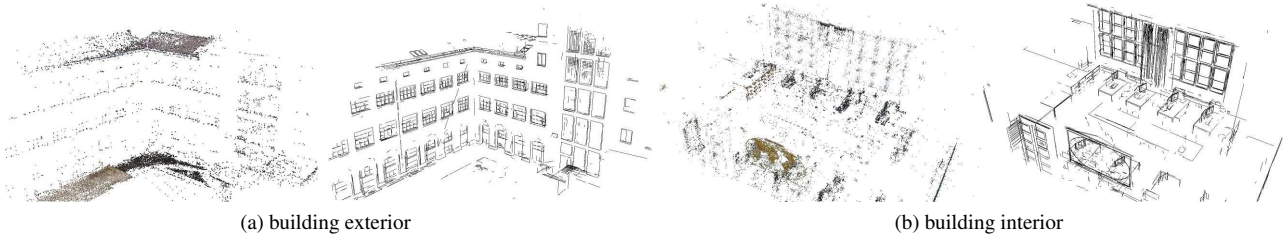


Figure 3: Sparse point cloud (left) and corresponding 3D line segments (right) for building exterior (a) and interior (b) of the *Office* dataset.

reconstruction step assumes images to be undistorted, radial distortion in the images should be removed in advance or modeled within the SfM process. Further, both models need to be approximately equally scaled. This can be achieved by fixing the scale in the SfM process by including one known real-world distance, the usage of GPS information, or a calibrated stereo camera configuration. Although the building interior often consists of poorly textured walls - which translates into problems during image matching due to the low number of matchable feature points - a feasible number of feature points for the pose estimation process should be found in most cases.

Subsequently, the computed camera orientations and undistorted images are used to generate 3D line segments following the *Line3D* method proposed by Hofer *et al.* [8]. Figure 3 shows a comparison of the sparse point cloud obtained from the SfM process and the 3D line segment reconstruction of the building in Fig. 1. It can be clearly seen that the derived sparse point clouds do not contain information in low textured areas, while reconstructed feature points at the façade and window frames only populate on corners and junctions. A detection of shared structures in both models

based on the point cloud seems to be unfeasible. MVS approaches help to increase the density of the point cloud, but still perform bad in poorly textured areas like walls or windows, as exemplary shown in Fig. 1. Additionally, the enormous number of obtained 3D points handicap an efficient analysis of the scene structure. However, the reconstructed 3D line segments contain much more geometric information of the scene, particularly in terms of interpreting façades and windows. Additionally, analyzing 3D lines can be done far more efficient by the drastically lower number of lines compared to the densified point cloud, as noted in Table 1. The alignment of both models by matching corresponding 3D line segments of window frames seems reasonable.

We do not assume prior information of the building structure, but expect that window frames can be dissembled to orthogonal and co-planar 3D lines. This allows us to first define possible window plane hypotheses and then to reduce the matching problem from 3D to 2D.

3.3. Window Plane Hypotheses Generation

This section describes the generation of possible window plane hypotheses which are further used to apply 2D matching and find corresponding 3D line segments in both models.

Vertical Alignment Like many man-made constructions, the interior and exterior of buildings mostly consist of planar horizontal and vertical surfaces. This allows us for making use of the Manhattan-world assumption and first identify dominant orthogonal orientations by computing orientation histograms of the 3D lines followed by aligning the estimated vertical axis of the model according to the vertical axis of the coordinate system with the obtained rotation matrix. A similar approach is proposed by Furukawa *et al.* in [7].

Line Filtering In order to reduce the computational overhead and increase the robustness of the method, subsampling of the 3D lines is performed by eliminating cluttered and skewed 3D lines which unlikely belong to window frames following the Manhattan-world assumption. The set of 3D

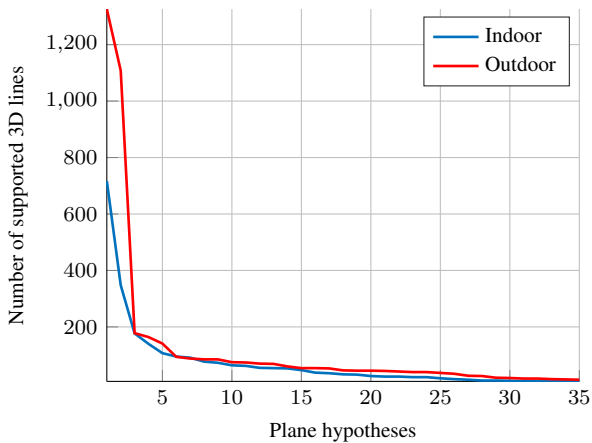


Figure 4: Number of supported 3D line segments for the 35 most dominant 3D planes in the *Office* dataset.

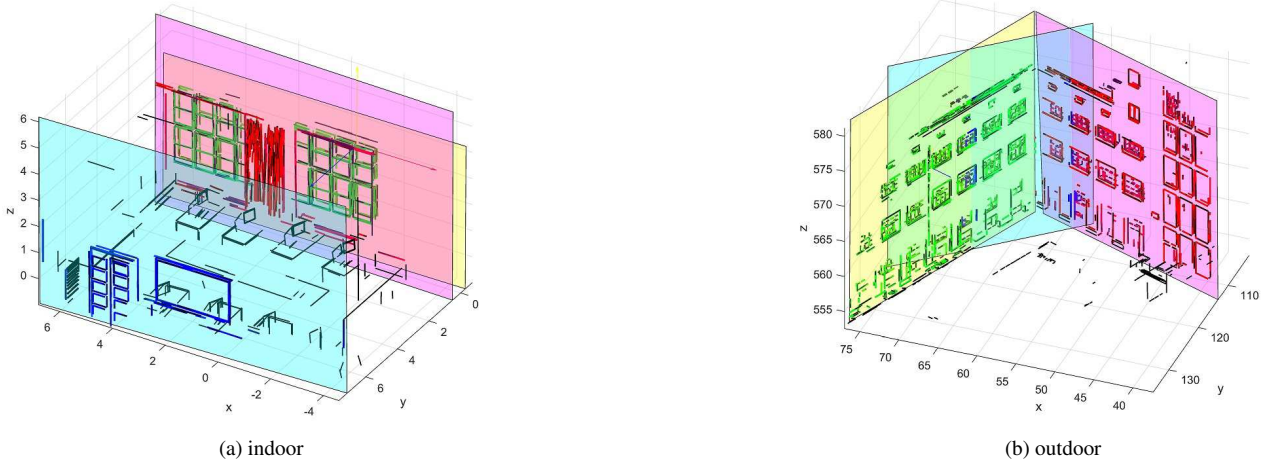


Figure 5: Filtered 3D lines and the three most dominant plane hypotheses in the *Office* dataset. Decreasing number of supporting 3D lines. (a) indoor: 717 (green), 348 (red), 177 (blue); (b) outdoor: 1326 (green), 1108 (red), 178 (blue).

lines $l \in L$ with length $|l|$ and vertical component of the normalized orientation φ_z are subsampled according to

$$L' = \{l \in L : |l| \geq \tau_l \wedge (|\varphi_z(l)| \leq \tau_\varphi \vee |\varphi_z(l)| \geq (1 - \tau_\varphi))\}, \quad (2)$$

where τ_l and τ_φ are user-defined thresholds defining a minimal length (e.g. 20 cm) and deviation along the vertical and horizontal axes (e.g. 0.05). Table 1 lists the number of remaining 3D line segments after the filtering step.

Plane Hypotheses From the set of remaining 3D line segments, multiple window plane hypotheses are generated by assuming co-planar window frames. A RANSAC estimation is applied to find dominant 3D planes, wherein inliers are identified as 3D lines lying on this plane within a threshold of the thickness of the plane. Each plane is defined by the intercept of close, orthogonal, and co-planar 3D line segments. The normal of the plane is directed towards the camera from which these lines were reconstructed in order to distinguish between indoor and outdoor sides.

We assume that window frames generate substantially more inliers compared to painted walls or other indoor and outdoor objects. Figure 4 plots the number of inliers for the first 35 generated 3D planes of the indoor and outdoor model shown in Fig. 1. As expected, the number of inliers decreases rapidly and only a few dominant 3D planes were found. Depending on the complexity of the building, it is mostly sufficient to consider the ten most dominant 3D planes. For the purpose of clarity, only the three most dominant 3D plane hypotheses together with their corresponding inlier lines are illustrated in Fig. 5.

For each pair of computed plane hypotheses, $\hat{\mathbf{T}}$ is now known up to a 2D translation vector within the outdoor plane.

The missing parameters can be estimated by first matching every plane hypothesis pair in 2D and then evaluating the matching result to find valid locations.

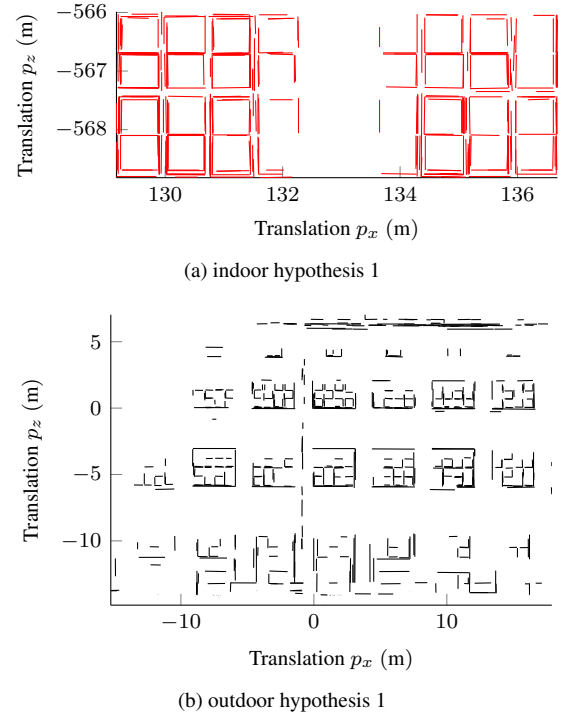


Figure 6: Projected 3D inliers onto the first plane hypothesis in (a) indoor and (b) outdoor scene.

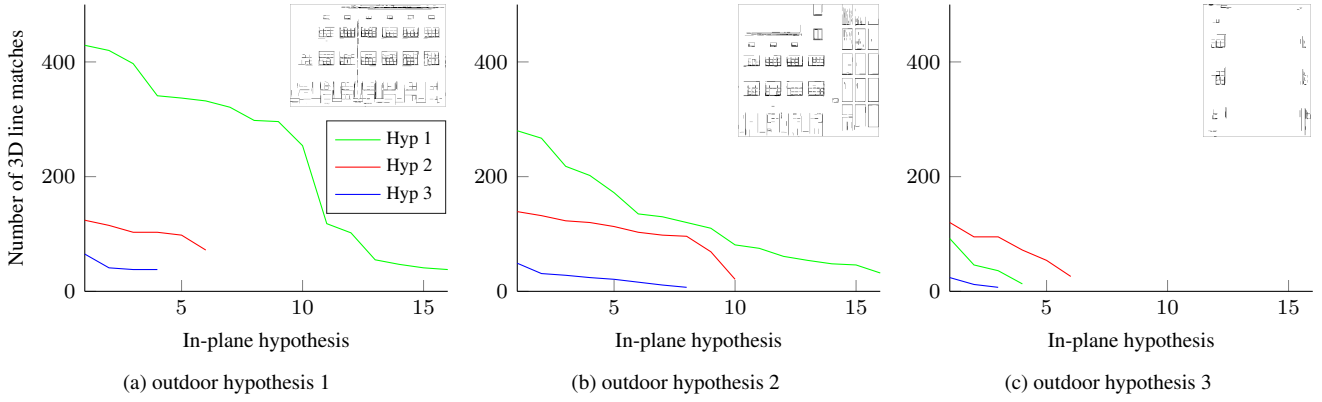


Figure 7: Number of 3D line matches for different in-plane hypotheses for the three most dominant indoor and outdoor hypotheses. (a-c) describe different outdoor hypotheses together with their corresponding 2D binary image. Different indoor hypotheses are indicated by different colors, while in-plane hypotheses are sorted by their number of matches.

3.4. Matching Plane Hypotheses

After computing multiple plane hypotheses, the next step is to determine corresponding plane hypotheses and find valid locations of the indoor model in the outdoor model in order to identify 3D line matches. This is done by performing oriented chamfer matching as described subsequently.

Binary Image Generation For each indoor and outdoor hypothesis, corresponding 3D lines considered as inliers by the plane estimations are projected onto their corresponding planes for generating 2D lines, as illustrated in Fig. 6. It has to be noted that, due to the reconstruction process, the models still contain inaccurate and missing lines, which has to be considered in the matching process. Furthermore, like most buildings, the façade shows highly repetitive structures. In this case, the correct location of the indoor model can not be identified without any further information like adja-

cent rooms. Instead, all possible valid locations should be returned by the method, whereby the correct one is identified by the user. As chamfer matching requires binary images, the 2D lines are discretized with a user-defined step size (*e.g.* 5 cm).

Oriented Chamfer Matching A popular and efficient technique for shape-based matching is provided by chamfer matching, particularly in presence of incompleteness and clutter. We make use of the oriented chamfer distance [12], which is defined as the mean distance of edge points of a template binary image to their closest edge points in a query binary image, weighted by the orientation differences of closest edge points. This distance can be efficiently computed using *distance transform*, while the orientations of the edge maps can be extracted directly from the 2D line segments.

The resulting chamfer distance map indicates possible locations of the indoor model, the so-called *in-plane hypothe-*

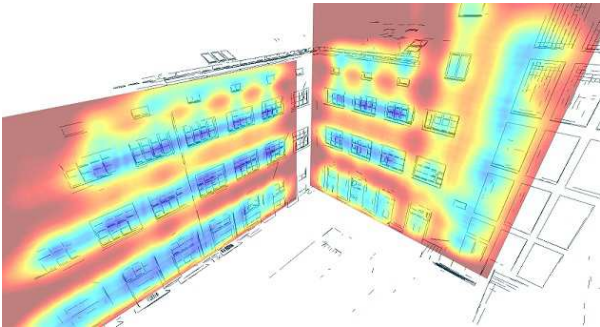


Figure 8: Chamfer distance maps of indoor hypothesis 1 and outdoor hypotheses 1 and 2 projected on outdoor 3D lines. Both maps are equally scaled, while blue color indicates low distance and therefore likely locations of the indoor model.

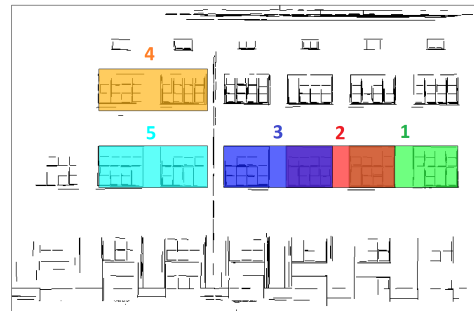


Figure 9: Most probable locations of the indoor model. The first five hypotheses belong to indoor plane hypothesis 1 and are all located on outdoor plane hypothesis 1.

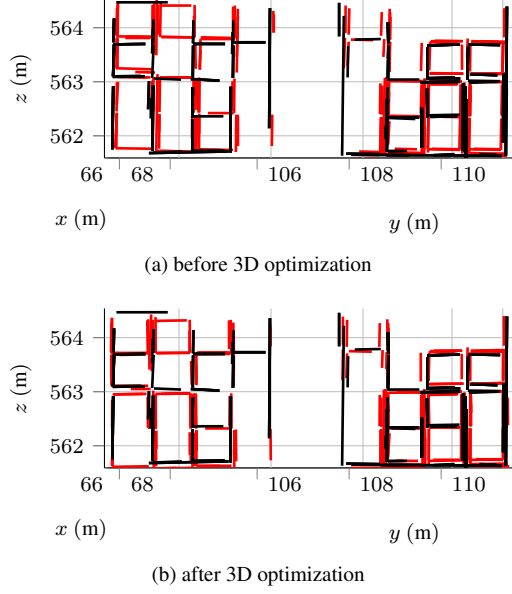


Figure 10: Refining the initial transformation by global 3D optimization: 429 3D line matches for (red) transformed indoor and (black) outdoor lines exemplary shown for the most supported hypothesis.

ses. Figure 8 illustrates the distance maps of matching indoor hypothesis 1 to outdoor hypotheses 1 and 2 projected onto the 3D lines of the outdoor model. Note the low distances for the windows at the first and second floor. However, differences in the scores for different floors are caused by missing edges during the reconstruction process and slightly different window heights for the first and second floor. Multiple in-plane hypotheses are subsequently identified by extracting local minima in the distance maps.

Finding Corresponding 3D Line Segments For each in-plane hypothesis i , a full initial transformation \hat{T}_i is now available. After transforming all indoor inlier 3D line segments with \hat{T}_i , corresponding 3D line segments can be detected as closest parallel 3D line segments of the outdoor model. Due to the plane estimation, discretization, and multiple window pane layer, the inlier 3D indoor lines are shifted along the normal orientation of the plane until a maximum number of matches is reached.

This procedure is repeated for all possible plane combinations and in-plane hypotheses, while the number of detected 3D line matches indicates the quality of the matching. Figure 7 shows the number of matches for each pair of planes and multiple in-plane hypotheses. Most matches are found by the correct indoor plane hypothesis 1 (green) and the first outdoor plane hypothesis (a), followed by the second façade (b), whereby numerous in-plane locations produce a similar number of matches. Wrong indoor plane hypotheses



(a) view from outside



(b) view from inside

Figure 11: Aligned point clouds of indoor and outdoor model from different perspectives.

(red and green) and the wrong outdoor plane hypothesis (c) generate significantly less matches. Figure 9 illustrates the location of the five most probable in-plane hypotheses. All of them correspond to the first indoor plane hypothesis and first outdoor plane hypothesis.

3.5. 3D Refinement

After obtaining the n most probable in-plane hypotheses and manually choosing the correct one, the parameters of the initial transformation \hat{T} are still erroneous caused by inaccurate plane estimations, the discretization, or unequal scale of both models, as exemplary shown in Fig. 10 (a).

A fine alignment is achieved by using the obtained 3D line matches and minimizing Eq. (1). Due to the fact, that corresponding 3D line segments still can vary in their distance - as they could be fragmented during the 3D line generation step - they are extended to infinity. Therefore, the perpendicular distance between matched lines is minimized. Note that this optimization requires both horizontal and vertical line matches in order to eliminate one degree of freedom, but should be satisfied in most cases.

Table 1 summarizes the intermediate results of the alignment and the effect of the global optimization. The mean of all perpendicular distances of 3D line matches can be considered as a measure of the alignment accuracy and results in 4.7 cm for the *Office* dataset. A visualization of the aligned 3D line matches before and after the global optimization is illustrated in Fig. 10, while Fig. 11 shows the final alignment of both dense point clouds.

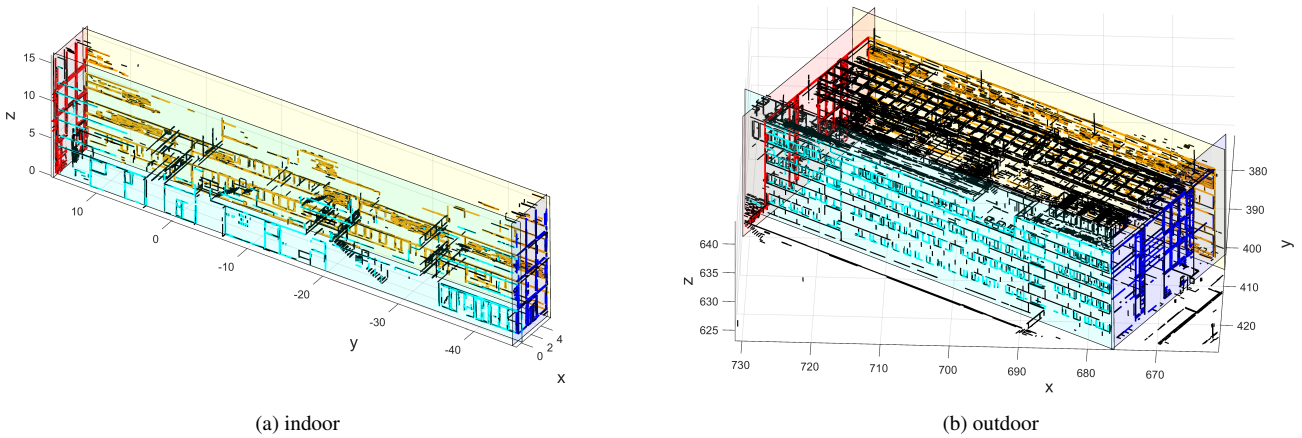


Figure 12: Filtered 3D lines and the four most dominant plane hypotheses in the *Building* dataset. Decreasing number of supporting 3D lines. (a) indoor: 2827 (orange), 982 (cyan), 412 (red), 272 (blue); (b) outdoor: 2395 (orange), 1617 (cyan), 721 (red), 676 (blue).

4. Experiments

Beside the dataset and result in the sections before, another experiment was carried out to illustrate the performance of the method. After giving an overview about the data acquisition and properties of the dataset, intermediate and final results of the alignment are described.

4.1. Dataset Description

The *Building* dataset contains an outdoor image sequence of a complete building captured from an *UAV* and an indoor hand-held image sequence inside of the building basement. Two large windows at both face sides of the building can be used for stitching the indoor and outdoor model. GPS tags of the aerial images were included in a SfM pipeline to compute a georeferenced, vertically aligned, and correctly scaled reconstruction model. However, one known real-world distance and direction has been included in the indoor reconstruction in order to approximate the orientation and scale of the indoor model. 3D line segments of both models were further generated using the *Line3D* method proposed in Section 3.2 (*cf.* Fig. 12). A description of the scene and intermediate results for this dataset are given in Table 1. For further information of this freely available dataset, please refer to [9].

4.2. Alignment Result

Unlike the dataset used in section 3, the alignment of these models is unique up to a 180° rotation of the indoor model, while the connection can be achieved on both windows sides of the building. The result of the plane hypotheses generation is illustrated in Fig. 12. The four most dominant plane hypotheses represent the four façades of the outdoor model

and the two walls and two window sides of the indoor model. In this dataset, the number of matchable plane hypotheses can be reduced by a bounding-box criteria. As the indoor model should not break through the outdoor model, the front and back sides of the indoor model are not matched to the side façades of the outdoor model. Further, as the side walls of the indoor model have no connection to the outdoor model

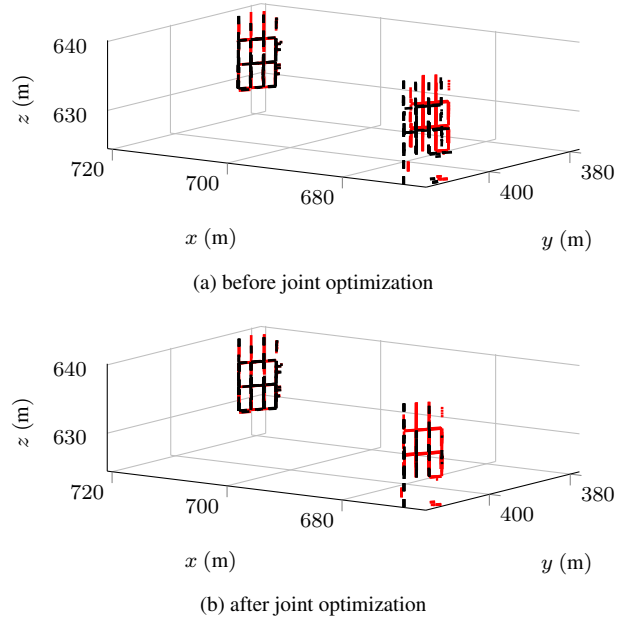


Figure 13: Visualization of (red) transformed indoor and (black) outdoor 3D line matches considering only matches at one face side of the building (a) and matches at both sides (b).

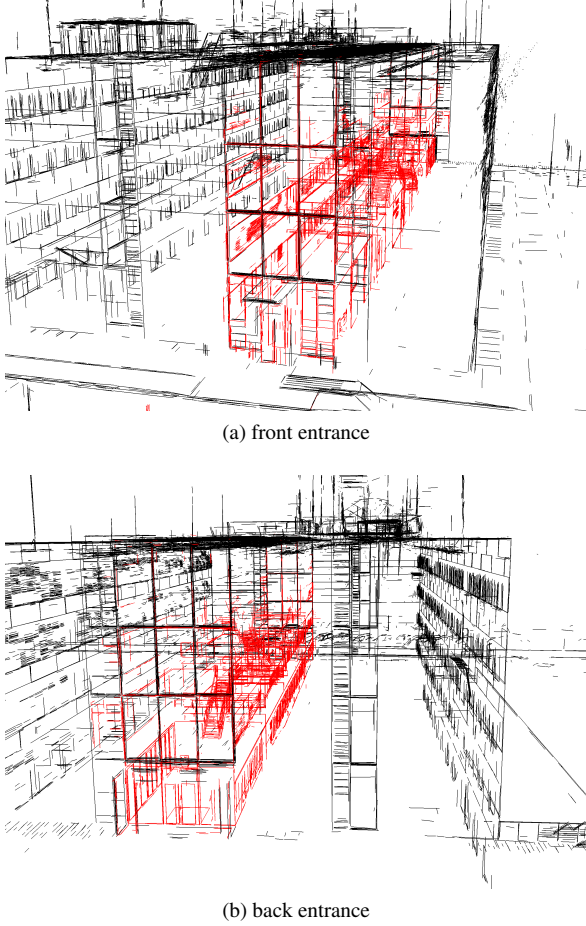


Figure 14: Final result after joint optimization of front and back entrance from different perspectives.

and contain different structures, only two main hypotheses remain after the 2D matching step.

172 inlier 3D matches were found when matching one window façade. Due to the building structure, another 157 3D line matches can be added when considering the second hypothesis on the opposite window façade, as shown in Fig. 13(a). If only matches at one side of the building are being used, small inaccuracies of the estimated rotation and scale together with the elongated structure of the building (60 m) cause an imprecise fit observed at the opposite side of the building (*cf.* Fig. 13(a)). Therefore, a joint optimization with matches at both sides is performed which leads to an accurate and robust estimation of \mathbf{T} with a mean error of 5.3 cm (*cf.* Figure 13(b)). Figure 14 shows the final alignment of all 3D lines viewed from both sides of the building.

5. Discussion and Future Work

We have presented an approach for automatically aligning individual indoor and outdoor reconstructions that uses

Table 1: Properties, intermediate and final results of the experiments *Office* and *Building*. Note the relatively small number of 3D line segments compared to the densified point clouds generated by a standard MVS [11]. Errors are defined as the mean perpendicular distance between 3D line matches before and after global optimization.

	Dataset			
	Office		Building	
	In	Out	In	Out
Base area (m ²)	75	405	360	1500
Images	247	41	320	228
3D Points (Mio)	9	18	13	134
3D lines	4373	3905	10315	23801
Filtered 3D lines	1724	2764	6616	21385
Matches	429		329	
Error before optimization (cm)	5.7		47.7	
Error after optimization (cm)	4.7		5.3	

SfM and a 3D line segment reconstruction algorithm. As connecting those kinds of models is mostly restricted to their geometric shapes like windows and doors, 3D lines are well suited for this task. Compared to the extensive generation and analysis of dense 3D points using *Multi-View Stereo*, a comparatively small number of 3D lines offer more interpretable information, at least in detecting and matching geometric shapes.

The proposed system exploits the planar structures of buildings for generating multiple meaningful matchable hypotheses and is therefore not limited by the complexity of the building. After detecting multiple 3D plane hypotheses, matching can be applied efficiently in 2D by binary image matching methods. However, a more discriminative matching method has to be developed for our task, as standard methods return too many local in-plane minima and hence result in too much computational overhead. This is also the case for reducing the number of meaningful plane hypotheses. A preceding labeling of the 3D line segments using semantic image segmentation could help to include useful priors in the window plane estimation and 2D matching steps.

Beside aligning indoor and outdoor models, this method can also be extended to align individual adjacent room models which are connected by doors. In case of complex building interiors containing multiple rooms, a graph-based approach has to be developed in order to find the correct room constellation.

References

- [1] C. Wu. Visualsfm: a visual structure from motion system. <http://ccwu.me/vsfm/>. Accessed: 2016-03-20.
- [2] The Chillon Project: Aerial/ Terrestrial and Indoor Integration. <https://support.pix4d.com/hc/en-us/articles/202557199-Scientific-White-Paper-The-Chillon-Project-Aerial-Terrestrial-and-Indoor-Integration>. Accessed: 2016-03-20.
- [3] Pix4d. <http://www.pix4d.com/>. Accessed: 2016-03-20.
- [4] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf. Parametric Correspondence and Chamfer Matching: Two new Techniques for Image Matching. Technical report, DTIC Document, 1977.
- [5] C.-A. Brust, S. Sickert, M. Simon, E. Rodner, and J. Denzler. Convolutional Patch Networks with Spatial Prior for Road Detection and Urban Scene Understanding. In *Proceedings of the IEEE International Conference on Computer Vision Theory and Applications (VISAPP)*, 2015.
- [6] A. Cohen, T. Sattler, and M. Pollefeys. Merging the Unmatchable: Stitching Visually Disconnected SfM Models. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2129–2137, 2015.
- [7] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Manhattan-World Stereo. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1422–1429, 2009.
- [8] M. Hofer, M. Maurer, and H. Bischof. Line3D: Efficient 3D Scene Abstraction for the Built Environment. *Pattern Recognition*, pages 237–248, 2015.
- [9] T. Koch, P. d’Angelo, F. Kurz, F. Fraundorfer, P. Reinartz, and M. Körner. The TUM-DLR Multimodal Earth Observation Evaluation Benchmark. In *Proceedings of the CVPR Workshop on Visual Analysis of Satellite to Street View Imagery (VASSI)*, 2016.
- [10] A. Martinovic, J. Knopp, H. Riemenschneider, and L. Van Gool. 3D All The Way: Semantic Segmentation of Urban Scenes From Start to End in 3D. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4456–4465, 2015.
- [11] M. Rothmel, K. Wenzel, D. Fritsch, and N. Haala. SURE: Photogrammetric Surface Reconstruction from Imagery. In *Proceedings of the LC3D Workshop*, volume 8, 2012.
- [12] J. Shotton, A. Blake, and R. Cipolla. Multiscale Categorical Object Recognition using Contour Fragments. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(7):1270–1281, 2008.
- [13] N. Snavely, S. M. Seitz, and R. Szeliski. Photo Tourism: Exploring Photo Collections in 3D. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846, 2006.